CEUR Workshop Proceedings 2018 vol.2260, pages 175-185

Services for formation of digital documents metadata in the formats of international science-based databases

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 CEUR-WS. All rights reserved. This paper contains the review of modern scientometric databases. The specificity of the representation of scientific materials in them is highlighted. The integration methods based on automation of the process of creating metadata for documents, included in digital scientific collections, are presented. Features of the formation of metadata for international scientometric databases on mathematical and computer sciences are noted. An algorithm for the automated formation of metadata in the format of the Russian scientific citation index (RSCI) is given. To automatically parse the text of articles, several regular expression patterns have been created, with the help of which the main metadata groups were selected. The algorithm is implemented as a service, consisting of modules for analyzing the structure of documents, automatically selecting documents according to the established order (for example, lexicographic), extracting the annotation block, the alphabetical index generating module, creating a bibliographic description of the article for writing headers of this article, converting documents to the portable document format (pdf), according to the determined parameters. The final module is the formation of metadata for exports to the RSCI. Approbation of the algorithm for the collection of articles of the journal “Russian Digital Libraries” was noted. The service for the formation of metadata for the documents of the digital collection Lobachevskii DML, made in accordance with the diagrams of the fundamental metadata of the European Digital Mathematical Library (EuDML) and the bibliographic database DBLP, is presented. Templates for showing metadata of articles of digital collections Lobachevskii DML in accordance with the scheme NISO JATS V1.0 are prepared. Plugins for the Open Journal System, allowing generation of metadata for science-based databases for downloadable articles are developed.

Keywords

Computer sciences, DBLP, Digital collections, EuDML, European Digital Mathematical Library, Mathematical, Metadata, Scientometric databases, Semantic methods, Structural, Stylistic analysis of digital documents

References

[1] Gerasimov A.N., Elizarov A.M., and Lipachev E.K. Subsystem of Formation Metadata for Science Index Databases on Management Platform Electronic Scientific Journals // Russian Digital Libraries Journal. – 2015. – No 1-2 (18). – P. 6–31.

[2] Gerasimov A.N., Elizarov A.M., Lipachev E.K., and Khaydarov S.M. Metody avtomati-zirovannogo izvlecheniya metadannykh nauchnykh publikatsiy dlya bibliografiche-skikh i referativnykh baz tsitirovaniya // Sb. nauch. statey XIX Ob"yedinennoy konferentsii «Internet i sovremennoye obshchestvo» IMS-2016. – SPb, 2016. – P. 41–48.

[3] Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM. Programma avtomatizirovannogo formirovaniya metadannykh v formate Rossiyskogo indeksa nauchnogo tsitirovaniya dlya statey zhurnala «Elektronnyye biblioteki» / A.M. Elizarov, E.K. Lipachev, S.M. Khaydarov; zayavitel' i pravoobladatel' KFU (RU). – №2017663206; decl. 19.12.2017; publ. 16.02.2018, Reyestr programm dlya EVM.

[4] Elizarov A.M., Lipachev E.K., and Khaydarov S.M. Automated System of Services for Processing of Large Collections of Scientific Documents // CEUR Workshop Proceedings. – 2016. – Vol. 1752. – P. 58–64.

[5] Sylwestrzak W., Borbinha J., Bouche T., Nowiński A., and Sojka P. EuDML – Towards the European Digital Mathematics Library // In: Sojka P. (ed.): Towards a Digital Mathematics Library. Paris, France, July 7–8th, 2010. Masaryk University Press, Brno, Czech Republic, 2010. – P. 11–26. URL: https://dml.cz/handle/10338.dmlcz/702569.

[6] Bouche T., Goutorbe C., Jorda J.-P., and Jost M. The EuDML Metadata Schema: Version 1.0 // In: Sojka P., Bouche T. (eds.): Towards a Digital Mathematics Library. Bertinoro, Italy, July 20–21st, 2011. Masaryk University Press, Brno, Czech Republic, 2011. P. 45–61. URL: https://dml.cz/handle/10338.dmlcz/702602.

[7] Bouche T., and Rákosník J. Report on the EuDML External Cooperation Model // In: Kaiser K., Krantz S., Wegner B. (Eds.): Topics and Issues in Electronic Publishing, JMM, Special Session, San Diego, January 2013. – P. 99–108. URL: https://www.emis.de/proceedings/TIEP2013/07bouche_rakosnik.pdf.

[8] Developing a 21st Century Global Library for Mathematics Research. Washington, D.C.: The National Academies Press, Washington, D.C (2014). arxiv.org/pdf/1404.1905, https://www.nap.edu/catalog/18619/developing-a-21st-century-global-library-for-mathematics-research

[9] Elizarov A.M., Lipachev E.K., and Zuev D.S. Digital Mathematical Libraries: Overview of Implementations and Content Management Services // CEUR Workshop Proceedings. – 2017. – Vol. 2022. – P. 317–325.

[10] Ion P.D.F., and Watt S.M. The Global Digital Mathematics Library and the International Mathematical Knowledge Trust // Geuvers H. et al. (Eds.): CICM 2017, LNAI 10383. 2017. – P. 56–69. https://doi.org/10.1007/978-3-3-9-62075-6 5.

[11] Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM. Programm-nyy kompleks vydeleniya metadannykh iz kollektsiy fiziko-matematicheskikh do-kumentov, predstavlennykh v formate OpenXML / A.M. Elizarov, E.K. Lipachev, S.M. Khaydarov; zayavitel' i pravoobladatel' KFU (RU). – №2016616511; decl. 21.06.16; publ. 08.08.16, Reyestr programm dlya EVM.

[12] EuDML metadata schema specification (v2.0 – final). URL: https://initiative.eudml.org/ eudml-metadat--schema-specification-v20-final.

[13] DBLP XML submission format guidelines. URL: http://dblp.uni-trier.de/faq/dblpsubmission.xsd.

[14] Elizarov A.M., and Lipachev E.K. Semanticheskiye metody i instrumenty elektronnoy matematicheskoy biblioteki Lobachevskii-DML // Nauchnyy servis v seti Internet: trudy XIX Vserossiyskoy nauchnoy konferentsii (18–23 september 2017, Novorossiysk). – Moscow: M.V. Keldysh Institute of Applied Mathematics, 2017. – P. 130–136. URL: http://keldysh.ru/abrau/2017/73.pdf. https://doi.org/10.20948/abrau-2017-73.

[15] Elizarov A.M., and Lipachev E.K. Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University // CEUR Workshop Proceedings. – 2017. – Vol. 2022. – P. 326–333.